DOCUMENT RESUME

ED 050 893 RE 003 436

AUTHOR Jongsma, Eugene R.
TITLE The Cloze Procedure: A Survey of the Research.
INSTITUTION Indiana Univ., Bloomington. School of Education.
NOTE 46p.; An Occasional Paper in Reading
AVAILABLE FROM The Reading Program, 222 Pine Hall, Indiana University, Bloomington, Indiana 47401 ($1.00)

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Cloze Procedure, *Literature Reviews, *Measurement Instruments, Readability, Reading Comprehension, *Reading Research, *Teaching Techniques

ABSTRACT

The purpose of this study is to survey research literature concerning the use of the cloze procedure and to determine what is known about the procedure. The survey proceeds in three broad areas: (1) methodological considerations, (2) cloze as a measuring device, and (3) cloze as a teaching technique. It is concluded that cloze based on every -n(th) deletion correlates more highly than based on selective deletions and that scoring of exact replacements is the most useful scoring system. The author reports that the cloze procedure is a valid and reliable measure of comprehension ability. He also feels that the most significant contribution the cloze procedure has made to reading research is in the aspect of readability, and this signals cloze as an important tool in the study of language. Further research is recommended to examine the construct validity of the cloze procedure and to investigate various uses of the cloze (1) to measure information gain, (2) to measure listening comprehension, and (3) to explore the relationship of factors such as literary style and attitude to comprehension. The author has found little research done on cloze procedure as a teaching technique, and this research evidence does not suggest the cloze as an effective teaching technique. (AW)

OCCASIONAL PAPERS IN READING
INDIANA UNIVERSITY SCHOOL OF EDUCATION

THE CLOZE PROCEDURE: A SURVEY OF THE RESEARCH

Eugene A. Jongsma
Indiana University

OCCASIONAL PAPERS IN READING

Occasional Papers in Reading is issued on a non-regular basis eight to twelve times per year by the Reading Program of Indiana University, Dr. Leo C. Fay, Chairman. Papers can deal with any aspect of reading, including articles describing research projects, review of research, theoretical speculation on reading, reports and evaluations on innovative program developments, review of tests and instructional materials, annotated or analytical bibliographies, and other general or specialized papers of current interest to the reading researcher or practitioner.

The length of the paper will vary depending upon the audience and the purpose for which the paper is written. Manuscripts should be submitted to the Editor, Occasional Papers in Reading, School of Education, Indiana University, 238 Pine Hall, Bloomington, Indiana, 47401. All submitted papers are evaluated by the Editorial Advisory Board: Dr. Carl B. Smith, Chairman, Dr. William Blanton, and Dr. James Laffey. Copies of papers in the series can be obtained for $1.00 each from the Reading Program office, 222 Pine Hall.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

THE CLOZE PROCEDURE: A SURVEY OF THE RESEARCH

Wilson L. Taylor is generally credited with being the "father of the cloze procedure." Completion-type exercises had been used previously in measures of intelligence and teacher-constructed tests. Deletions in these tests were usually highly selective and focused on high-content words that conveyed meaning. However, the cloze procedure, as introduced by Taylor (1953), required the systematic, mechanical deletion of words. In constructing the cloze procedure, Taylor drew upon Miller's work in communication theory, Osgood's "dispositional mechanisms," and the principles of statistical random sampling. His definition of cloze, which has also been accepted by most others working with the cloze procedure, considers cloze "a method of intercepting a message from a 'transmitter' (writer or speaker), mutilating its language patterns by deleting parts, and so administering it to 'receivers' (readers and listeners) that their attempts to make the patterns whole again potentially yield a considerable number of cloze units." (1953, p. 416)

The purpose of this paper is to review the literature concerning the use of the cloze procedure, to organize, analyze, and synthesize this literature in order to determine what is now known about the cloze procedure, and to raise unanswered questions that could serve to guide future research. The review of the literature will be grouped into three broad areas: (1) methodological considerations; (2) cloze as a measuring device; and (3) cloze as a teaching technique.

Cloze Methodology

Instrument Construction

Perhaps it is misleading to use the term "cloze procedure," for there has been no one universally accepted procedure for constructing cloze exercises. Some of the problems facing the developers of cloze instruments have been determining the appropriate length of cloze passages, the type of deletions to be made, the rate of deletions, and the total number of deletions per passage.

A number of investigators have examined the problems of instrument construction. In his early work, Taylor (1953) found that every-fifth deletions were successful in measuring readability, providing there were more than 16 cloze blanks per passage. He seems to have arrived at the every-fifth deletion rate rather arbitrarily, as there was little research upon which to base that judgment at that time. In later studies (1956, 1957) he assessed the effects of selective deletions of "easy" (structural) words and "hard" (semantic) words. He maintained his earlier conclusion that every-fifth word deletions were most effective, especially in measuring the relative difficulty of different materials but increased the suggested test length to 50 cloze items to insure a more representative sample. An interesting aspect of Taylor's work was the discovery that cloze tests based on structural deletions were measuring different sorts of things than cloze tests based on semantic deletions. This issue was later explored by a number of researchers.

In reviewing the validity and utility of the cloze procedure, Rankin (1959b) adapted Fries' division of language in describing the effects of selective deletions. Rankin contended that "lexical comprehension,"

measured through the deletion of only nouns and verbs, involves the understanding of substantive content while "structural meaning," assessed by deleting function words, such as articles, is an understanding of the interrelationship of ideas. Rankin pointed out that research up to that point had suggested that the every-n^{th} deletion system correlated more highly with intelligence, while the selective deletion system, based on form class, was more closely related to the subject's knowledge of the content of the passage.

A study by MacGinitie (1961) lent further support to a deletion rate of every-fifth word. He used 15 different omission sets representing eight patterns to investigate constraints within complete prose paragraphs differing in content, style, and difficulty. His results show "...that additional uninterrupted context beyond five words did not help in the restoration of the missing words." (p. 127)

In another study comparing rates of deletion, Fillenbaum (1963) found that the number of successful responses increases moderately with decreases in rate of deletions. Unlike MacGinitie, however, Fillenbaum did not clearly define at what point increased context fails to produce a corresponding increase in predictability. The tests used in this study were long (minimum of 200 deletions each), and this may have had an additional effect on the subjects' ability to produce successful responses.

Bormuth (1964a) examined the effects of different cloze forms particularly with respect to their reliability in measuring the comprehension difficulties of passages. Keeping the rate of deletion constant (every-fifth word), he varied the starting point to produce the five possible forms for each of 20 passages. After administering different forms to

each of five matched groups, differences were found between the means of the various forms. Bormuth concluded that one should not use only a single cloze test form over a given passage, especially where precise determination of passage difficulties are needed. The results also indicated that cloze exercises of less than 50 items tend to be unreliable; this is consistent with Taylor's earlier findings.

Others (Greene, 1965; Heitzman and Bloomer, 1967; Weaver and Bickley, 1967) have explored the effects of selective deletions as contrasted with every-n^{th} deletions. Greene (1965), for example, found that a modified cloze procedure based on rational deletions of nouns, verbs, adjectives, and adverbs, produced a better test in terms of reliability and item performance. The difficulty of the modified cloze test was distributed more evenly over all the items than was the difficulty of the standard cloze test (every-twelfth deletions), which had more variation in item difficulty. One wonders what Green's rationale was for using an every-twelfth deletion system, when much of the previous research had substantiated an every-fifth deletion rate.

Little attention has been directed to some aspects of constructing cloze tests. Many investigators have expressed confusion over just what constitutes a deletion element. For example, should numerals be subject to deletion? Should hyphenated words be counted as single words or divided into their separate parts? These questions are usually answered on a logical basis by individual researchers. No research evidence is available for guidance.

Test Administration

Most of the investigators using the cloze procedure have presented the subjects with cloze exercises based on passages which they, presumably, have never seen before. A few exceptions to this practice have been tried.

Coleman (1962) allowed his subjects 50 seconds to read the unmutilated passage before taking a cloze test over the same passage. This procedure would appear to change the nature of the task, for example, by increasing the influence of short term memory.

The other variation in administration involves using the cloze procedure to measure "information gain," or the amount of knowledge actually due to reading the passage. In two studies done with Air Force Trainees, Taylor (1956, 1957) used a "before" cloze test, allowed one week for study of the article, and then administered the "after" cloze test. These "after" minus "before" results were a highly significant measure of learning. Similar procedures have been used by Rankin (1957) and Bormuth (1969b).

In general, very little research has been done on the aspects of cloze test administration. No one has considered the effects of various types of directions or what influence an introductory sample exercise may have. How does the pre-reading of a passage change the nature of the task? Should subjects be encouraged to guess? These are some areas that deserve further exploration.

Scoring

Probably the most widely researched aspect of cloze methodology has been that of scoring. The prevailing rule-of-thumb in scoring cloze

test is to count correct only the exact replacement of the deleted word. Much of the research has focused on the benefits of scoring synonyms as compared with exact replacement. The results have been fairly consistent.

Taylor (1953) found that scoring synonyms was not worth the extra effort. Ruddell's (1964) results showed no significant difference between the two scoring systems, synonym count and exact replacement, in terms of validity and reliability. The one exception was significantly higher reliability using a synonym count on passages using extremely high frequency patterns of language structure. Gallant (1964) compared exact replacement with a "substitute score." "Substitutes" consisted of responses which approximated, to a reasonable extent, the meaning of the word deleted and agreed in person and tense. Although slightly higher correlations were obtained between the cloze tests and the standardized tests when using the substitute scoring systems, the difference was not statistically significant. On the contrary, "the use of substitute scores...decreased the efficiency and the objectivity of the scoring procedure." (p. 58) Miller and Coleman (1967) used a weighted scoring system which included three points for exact replacements, two points for synonyms, and one point for the correct word class. Since the weighted score correlated .99 with the exact replacement score, the authors concluded that weighted scoring was not worth the tim , unless investigators were particularly interested in synonyms.

Studies by Bormuth (1965a) and Fillenbaum (1963) categorized cloze responses into grammatical and lexical classes. In analyzing his results, Bormuth found that scores based on grammatically correct responses correlated positively with the criterion measure of comprehension ability, while scores based on grammatically incorrect responses

either correlated negatively or not at all with the criterion measure. Furthermore, "among just the grammatically correct response categories, the correlations with the criterion increased as a function of the similarity of the meanings of the responses to the deleted words." (p. 284) Through multiple regression analysis, Bormuth found that scores based on exact replacements accounted for about 95 per cent of all the variance in the entire set of cloze tests. Fillenbaum went a step further and analyzed responses according to the traditional parts of speech. His analysis indicated considerable differences in performance both between and within grammatical classes.

Other rather unusual variations in scoring were examined by Hafner (1964) and Musgrave (1963). Hafner used two scoring procedures. The first was a ratio of the percentage of connective word completions correct to the percentage of content word completions correct. In effect, this gave a comparison of the structural comprehension with the lexical comprehension, concepts mentioned previously in this paper. His second procedure, the G.C.I.A. score, measured the percentage of responses correct grammatically of those cloze responses that were incorrect (not exact replacements). The G.C.I.A. score correlated higher with criterion measures of intelligence and vocabulary. Musgrave compared exact replacements with "commonality" scores. In the commonality procedure, she counted correct those responses which were exact duplicates of the most popular responses made by her group of subjects. Results showed the exact replacement score to be highly correlated with the commonality score.

Thus it appears that the literature consistently shows the scoring of exact replacements to be the most objective, efficient, and useful scoring system to use with the cloze procedure. Although slightly higher reliability

has been obtained at times using other procedures, such as a synonym count, the increased time and subjectivity necessary for such systems do not warrant their use. The exception to synonym usage may be in using the cloze procedure as a teaching technique. This usage will be discussed in a later section of this paper.

In summary, the mechanical, every-fifth word deletion rate, has been the most widely used and accepted type of cloze procedure. Other investigators have used selective deletions of nouns, verbs, and other specific elements for particular purposes. The research has suggested that cloze tests based on the deletion of structural words are measuring the interrelationships of ideas and correlate more highly with intelligence measures. Cloze tests based on deletions of nouns and verbs are a better measure of a subject's knowledge of the content of a passage and do not correlate as highly with intelligence. The every-fifth deletion approach assumes that because of semi-random sampling, a representative number of structural elements, as well as lexical elements, will be deleted in each passage. Scoring exact replacements has proved to the most objective and useful scoring procedure. Cloze tests are usually administered to subjects who have not read the passages before and are given in an untimed situation.

Cloze As A Measuring Device

Comprehension Ability

Much of the research has focused on the validity and reliability of the cloze procedure as a measure of comprehension ability. Early studies, quite appropriately, were specifically interested in determining the validity and reliability of cloze in this area. Later, studies used cloze for more specific purposes, yet often added further evidence to the validity and reliability of the procedure. The following sections

will review the literature related to different aspects of comprehension.

Cloze as a measure of specific comprehension. Various investigators have examined the validity of the cloze procedure by correlating the results of cloze tests with other measures of comprehension, usually multiple-choice questions, over the same material. In using this approach it is important to consider the nature of the comprehension questions, how they were constructed, and, indeed, what skills they were measuring.

In two similar studies with Air Force trainees, Taylor (1956, 1957) administered pre and post cloze tests as well as pre and post multiple-choice comprehension tests over the same article. He allowed an interval of one week for trainees to study the article. Correlations between the cloze test results and the multiple-choice results ranged from .51 to .92. A major weakness in this study is Taylor's neglect to describe the comprehension questions he used. Did they measure literal details, main idea, or inferences? The reader is unsure of just what the cloze results have been correlated with.

Bormuth (1962, 1963) carefully constructed 31 multiple-choice items for each of nine passages, representing three different content areas. The items sampled seven comprehension skills--vocabulary, facts, sequence, relationships, main idea, inferences, and author's purpose. An equal number of items for each skill was included for each passage. The author also used a pilot study to validate the comprehension tests. Correlations between the cloze tests results and the multiple-choice responses ranged from .73 to .84. Bormuth also found the cloze tests to be highly reliable, with coefficients of .84 to .88.

Seven years later, Bormuth (1969a) replicated this study but went a step further to do a factor analysis of the results. The results again show cloze tests correlate highly with a variety of multiple-choice comprehension questions over the same passage. Under the condition of this study, Bormuth concludes th" cloze tests...measure skills closely related or identical to those me ired by conventional multiple-choice reading comprehension tests." (p. 365)

Ransom (1968) reported a study comparing the cloze procedure with an Informal Reading Inventory for students in first through sixth grade. Both the cloze tests and the Informal Reading Inventory were based on basal reader material. The correlations between the independent, instructional, and frustrational reading levels derived by the cloze tests with those derived by the Informal Reading Inventory were significant at the .01 level. The one exception was the first grade results where scores did not correlate as highly. The criteria for reading levels, using the cloze tests (independent above 50%, instructional above 30%, frustrational below 20%) were arbitrarily set with no convincing rationale or evidence from previous research.

Generally, studies have shown that the cloze precedure is a valid and reliable measure of specific comprehension. In correlating cloze results with other measures of comprehension some investigators have not made a point of carefully defining their "other" measures of comprehension and validating them. In such cases, correlations are meaningless.

Cloze as a measure of general comprehension. Other efforts have been made to validate the cloze procedure by correlating cloze results

with those of standardized reading tests. These efforts differ with those mentioned in the previous section which correlated cloze results with comprehension measures of the same passage.

Jenkinson (1957) was one of the first investigators to correlate cloze tests results with a standardized reading test. In constructing her cloze tests, she made selective deletions equally distributed over "easy," "more difficult," and "hard" words. The rate of deletion varied from every third word to every ninth or tenth word. She found that for her sample of 210 high school students, the cloze test results correlated significantly with the scores of the Cooperative Reading Test. Cloze results correlated .78 with the vocabulary section and .73 with the comprehension section.

In doing a factor analysis of the cloze procedure and other related measures, Weaver and Kingston (1963) found correlations between their cloze tests results and scores on the Davis Reading Test ranged from .25 to .51. They constructed their cloze tests by making either structural deletions or lexical deletions of both essays and speeches. The fact that deletions were not evenly distributed over "easy" and "hard" words may account, in part, for the low correlations obtained.

Hafner (1964) correlated cloze test results with performance on the Michigan Vocabulary Test for college students in a reading methods course and obtained a coefficient of .56.

Ruddell (1965a, 1965b) used an every-fifth word deletion system and controlled for high and low frequency patterns of oral language in constructing his cloze tests. Correlations with the paragraph meaning section of the Stanford Achievement Test ranged from .61 to .78. In addition, split-half reliability went as high as .97.

A somewhat different approach to validation was used by Gallant (1964, 1965) who used one of the alternate forms of the paragraph reading section of the Metropolitan Achievement Test in constructing a cloze test of every-fifth deletions for primary level students. Assuming that the cloze process would be too complex for first graders, she modified the procedure by offering three choices for each cloze unit, thus making it a multiple-choice task. This modification departs considerable from the theory upon which the cloze procedure is based. Furthermore, Gallant never gives a rationale for selecting the distractors in these multiple-choice items. Careful analysis reveals that in many cases, only two of the three choices are feasible, that is, would fit into the language pattern of that particular slot. Perhaps form class should have been considered in the selection of distractors. Students in grades two and three received the conventional, every-fifth word deletions, cloze tests. Cloze scores correlated from .65 to .81 with the standardized test results. Reliability of the cloze tests was high, ranging from .90 to .97.

Greene(1965) compared the results of his modified cloze precedure, mentioned earlier in this paper, with the Diagnostic Reading Survey. A correlation coefficient of .51 was obtained between cloze and total comprehension.

Although the literature is mixed, much of the research has demonstrated that the cloze procedure correlates substantially with standardized measures of reading comprehension. It appears that cloze tests using an every-n^{th}, usually an every-fifth, deletion system correlate more highly than cloze tests based on selective deletions of structural and lexical

elements. Most efforts at validation have been in the area of concurrent validity, that is, attempting to show a high degree of relationship between cloze test results and commonly accepted measures of comprehension. The following section will turn to construct validity and a look at the underlying processes involved in the cloze procedure.

Factor analysis of the cloze procedure. Weaver and Kingston (1963) conducted a factor analysis study of the cloze procedure to determine the proportions of variance which could be assigned to factors basic to more commonly used tests of vocabulary, language aptitude, and reading ability. They administered a series of standardized tests, as well as four reading cloze tests and four listening cloze tests to 160 college juniors. Correlations were determined and an attempt was made to isolate factors by means of orthogonal factor analysis. Three factors were identified: (1) a verbal comprehension factor; (2) a cloze factor; and (3) a rote memory, flexible retrieval factor. All cloze tests loaded moderately to high on the cloze factor, regardless of the type of deletions made, and regardless of reading or listening. The data suggested that cloze tests are more related to each other than to the other factors identified in this study. Contrary to most of the previous literature, Weaver and Kingston concluded that "cloze tests are related only moderately to the verbal comprehension factor." (p. 259)

Bormuth has reacted to Weaver and Kingston's study at various times (Bormuth and MacDonald, 1965; Bormuth, 1969a; Bormuth, 1969b). He warns that we cannot apply their conclusions to all cloze tests, especially those employing every-n^{th} deletions, because they only included selective deletions of structural and lexical words. Other criticisms which Bormuth

raises include: (1) subjects were highly select, i.e. college students; (2) the standardized tests showed unusual patterns of factor loadings; and (3) the cloze tests showed inconsistencies in their loading patterns.

A few years after the Weaver-Kingston study, Bormuth (1969a) conducted a factor analysis of cloze tests. Many differences exist between the two studies. Weaver and Kingston made selective deletions using essay and speech material; Bormuth deleted every-fifth word in materials from three different content areas. Weaver and Kingston's subjects were college juniors; Bormuth randomly selected fourth, fifth and sixth graders. Weaver and Kingston used standardized tests as criterion measures; Bormuth constructed his own multiple-choice comprehension questions and pre-validated them by professional judgment and pilot testing. Bormuth concluded from his study that "one factor accounted for the reponderance of the variance... little difficulty applying the name 'reading comprehension ability' to that factor." (p. 364)

One can see that the evidence is conflicting. There is no conclusive research on the construct validity of the cloze procedure. The fact is that the processes one must go through in completing a cloze test are relatively unknown. If one accepts the high positive relationships between cloze tests and tests of reading comprehension, perhaps the identification of the processes underlying cloze is closely tied with the processes of comprehension, itself. Jenkinson (1957) has probably come the closest of anyone to examining this question. She used the cloze procedure as a tool for getting at the product and process of comprehension. Using retrospective and introspective techniques in individual interview situations, she had high school subjects "think aloud" while responding to cloze tests. She was able to develop a classification system for analyzing the "process" and identified characteristics of good and poor readers. Perhaps what is

needed are more efforts like that of Jenkinson, as well as new and innovative ways to get at the underlying structure of comprehension.

Cloze as related to literary style. Literary style is another aspect of comprehension that has been related to the cloze procedure. Bormuth and MacDonald (1965) investigated the correlation of cloze test scores with scores on tests to measure the ability to detect an author's literary style. The investigators carefully developed tests to detect the literary style of two authors. In addition, they constructed cloze tests, using every-fifth deletions, based on works of the authors. A pre-cloze test was given to 150 female college students at the beginning of a literature course. After reading and studying the authors' works, the students took the post-cloze test and the tests to detect literary style. Because the two tests correlated highly, the investigators suggest that a person's sensitivity to literary style is one of the variables which effects performance on cloze tests. However, the extended length of the cloze tests may have been at play here. Most researchers have used passages of 250 words with 50 cloze items. The cloze tests in this study were 1000 words in length. Surely, this increased length would allow a student a much better chance for getting a taste of the authors' styles.

It is dangerous to draw many conclusions or attempt to generalize on the basis of one study. However, the investigators have raised some interesting questions. Literary style has particular relevance to readability; however, it is too often a neglected factor. Perhaps future research can aid in determining what effect an author's literary style has upon the reader's comprehension.

Cloze and affective correlates. A few investigators have begun to explore affective factors that effect cloze test performance. Others have used cloze as a tool in examining the relationship of affective variables with reading comprehension. The following studies will illustrate these efforts.

Manis and Dawes (1961) tested the hypothesis that readers who disagree with the contents of a controversial statement will be relatively insensitive to the writer's redundancy. They first gave their subjects, psychology students at the University of Pittsburgh, a semantic differential to assess attitudes toward the topic of capital punishment. They then constructed cloze tests over two passages, one favoring capital punishment and one opposing it. Students took both tests, and the results indicated that they performed more effectively on the passage that represented their own views. Manis and Dawes concluded that "cognitive performance may be adversely affected when the individual is presented with materials that contradict his beliefs." (p. 84)

Weaver, White and Kingston (1968) attempted to examine the interrelationships between reading comprehension, the reader's perception of himself, and his perception of a protagonist in a story. Semantic differentials and Cattell's 16 Personality Factor Questionnaire were used to obtain judgments of self and protagonist. Cloze results gave a measure of comprehension. Early results seemed to indicate a number of relationships between the cloze procedure and the affective measures. The affective component seemed to be more specific than is generally assumed. The investigators recognize that they are just beginning to explore this area and much more needs to be done.

Efforts like those mentioned above are initial attempts to examine areas that have not been studied in depth before. The point is, that the cloze procedure may serve as a useful tool in exploring the affective dimensions of comprehension.

Cloze as a measure of information gain. The term "information gain" is often used to describe the increase in knowledge actually due to the process of reading a passage. Many critics claim that conventional reading comprehension tests do not distinguish between a student's prior knowledge and the information he has gained from reading the test article. Information gain represents an attempt to get at this difference. The following studies have employed the cloze procedure in some fashion in determining information gain.

The methods used to measure information gain have varied. Taylor (1957) and Rankin (1959b) used similar procedures. They correlated measures of pre-cloze, pre-reading knowledge with post-cloze and post-reading knowledge. They found that cloze tests constructed by deleting selective nouns and verbs were more suitable at estimating pre-knowledge of the content. Both investigators found significant gains between pre and post-reading cloze tests.

Bormuth (1969b) matched pairs of students on the basis of a pre-cloze test and determined the appropriate level of difficulty of the reading passage to be used by administering a cloze readability test to one of each pair. Information gain was measured by subtracting the students' scores on a pre-reading multiple-choice test from their scores on the same multiple-choice test administered after reading the passage. According to Bormuth, the results showed "that scores on cloze tests do not depend solely upon a subject's

prior knowledge of the content of a passage. . .because as cloze scores increased, information gain increased." (p. 721)

Coleman and Miller (1968) used another means of determining information gain in addition to the usual post-cloze minus pre-cloze scores of matched subjects. The investigators employed Shannon's "guessing game technique" by having subjects guess each word of a passage in order. They recorded the number of guesses attempted and the number correct; then they repeated the same procedure using the same passage. Information gain equaled the increased number of correct guesses during the second reading. They found this procedure to be more effective.

It appears from the studies in this area that pre-cloze tests can be used to measure an individual's pre-knowledge of the content of a passage. Cloze tests have also been used to measure gains due to reading for groups of students. However, as Rankin (1964) has pointed out, when cloze tests are used to measure gains for individuals, regression effects must be taken into account just as on standardized reading tests. More research is needed to fully examine the merits of various measures of information gain. Studies in this area should aid in selecting the proper level of difficulty of materials which is likely to produce the greatest student gain.

Determining comparable criterion scores. The problem of interpreting cloze test results has plagued investigators ever since the cloze procedure has been used to measure comprehension ability. Some investigators have used the number of correct responses (raw score) in interpreting cloze test results. However, a raw score only has meaning for that particular cloze test, since tests vary as to length and number of cloze units.

Others have converted raw scores to percentages. While this allows the comparison of different cloze tests, it has little meaning in comparing cloze test results with conventional measure of reading assessment. The following studies represent efforts to build a framework in which to interpret cloze results in relation to more commonly used measures.

Bormuth (1967a) matched cloze test results with performance on multiple-choice comprehension tests of the same passage. He accepted the traditionally used criterion that comprehension scores between 75 - 90 per cent represent a student's instructional reading level and scores above 90 per cent indicate an independent reading level. His comparisons showed that a cloze test score of 38 per cent was comparable to a multiple-choice comprehension score of 75 per cent and that a cloze score of 50 per cent was comparable to a comprehension score of 90 per cent.

In a later study, Bormuth (1968a) used the Gray Oral Reading Tests paragraphs for determining cloze scores comparable to criterion comprehension and word recognition scores. He accepted word recognition scores of 95 per cent and 98 per cent as indicators of instructional and indenpendent levels. The results suggested cloze scores of 44 per cent and 57 per cent are comparable to comprehension scores of 75 per cent and 90 per cent. Bormuth attributes the six point difference between these results and those of the previously cited study to a ceiling effect on the multiple-choice test scores in the 1967 study. He also found cloze scores of 34 per cent and 54 per cent to be comparable to word recognition scores of 95 per cent and 98 per cent. The big discrepancy between the comprehension and word recognition cloze scores at the instructional level raises serious doubts about a 95 per cent word recognition score being comparable to a 75 per

cent comprehension score, says Bormuth. However, this discrepancy could be due to a number of other factors as well, such as incorrect scoring of oral reading errors and the emphasis of the subjects' previous reading instruction.

Ransom (1969) arbitrarily set cloze score criteria to represent independent, instructional, and frustrational levels and then merely correlated her cloze system with the results of an Informal Reading Inventory. Because correlations were high, she assumed her comparable cloze scores to be valid estimates of students' reading levels from second through sixth grades. In her system, cloze scores of 50 per cent or above represented the independent reading level; scores of 30 to 50 per cent represented the instructional level; and cloze scores below 20 per cent were at the frustrational level.

The studies above are honest attempts at building a framework to aid in making value judgments about cloze test results. However, they have serious limitations. They all are based on the notion of independent, instructional and frustrational reading levels and the commonly used percentages for defining such levels. The research that the definitions of these levels is based upon is far from convincing or conclusive. Any attempt to interpret cloze test results on the basis of such rationale must be treated cautiously. A more fruitful approach to the issue might be efforts like Bormuth's (1969a), with information gain, mentioned earlier, to determine the level of difficulty at which a student can profit most from material.

Cloze as a measure of listening comprehension. The cloze procedure has also been used to measure a subject's understanding of orally presented material. Taylor (1956) compared two radio scripts--one poorly written and the other well written. An announcer read the scripts but pushed a buzzer

and "thought" every fourteenth word. Subjects responded by writing down what they thought the missing word was. The overall results showed that the subjects performed better on the "good" script than on the "poorly written" one.

Weaver and Kingston (1963) used listening cloze tests in their factor analysis of the cloze procedure. Four listening tests were constructed by making either lexical or structural deletions of both essay and speech material. Although the authors reported that the listening scores loaded heavily on the "verbal comprehension" factor, zero order correlations between the cloze listening tests and the STEP Listening Comprehension test were relatively low (coefficients ranged from .45 to .54). This seems to suggest that the cloze listening tests were measuring factors other than listening comprehension.

Very little research has been done in this area so not a great deal of knowledge is known about the relationship of the cloze procedure to listening comprehension. Logical analysis would tell us that the process one goes through in performing a listening cloze test is different that that used in a reading cloze test. In taking a reading cloze test a subject is free to regress to the preceding context, try out new words in the context, and even look ahead in the context, all at his own pace. In a cloze listening test, it would seem that the subject would be able to mentally review only the preceding context to the extent that he could remember it. He would also be constrained to work the task at the same pace as the passage was being read orally. More research is needed to explore the use of cloze in measuring listening comprehension.

To summarize the foregoing research, it appears that the cloze procedure is a valid and reliable measure of both specific and general comprehension

ability. Independent studies using various levels of subjects have shown that, generally, cloze test results correlate highly with multiple-choice comprehension measures of the same passages and with standardized reading tests. A semi-random deletion system, such as every-fifth word, is more effective, although selective deletions of structural and/or lexical words are useful for specific purposes. Although concurrent validity has been demonstrated, there is little evidence as to the construct validity of the cloze procedure. Efforts are needed to define and examine the underlying processes involved in cloze which will consequently lead to a clearer conception of the components of comprehension.

The cloze procedure has been used to measure students' pre-reading knowledge of a passage and information gained from reading a passage. Research in these areas has not been as extensive or complete as in other areas. Future efforts might explore new approaches to measuring information gain, apart from the usual post-score minus pre-score procedure.

There is some evidence that cloze can be used to explore the relationship of factors such as literary style and attitude to comprehension.

Readability

Perhaps the most significant contribution the cloze procedure has made to reading research has come in the area of readability. Reviewers such as Chall (1958) and Klare (1963) have defined three major aspects of readability: (1) legibility; (2) interest; and (3) ease of understanding of comprehensibility. Cloze research related to readability has dealt almost entirely with the third aspect, the subject's ability to comprehend material due to certain language variables. The following section will

review various studies that have used the cloze procedure for readability purposes.

In the early stages of cloze research, investigators were interested merely in using the cloze procedure to rank passages according to the degree of difficulty. Taylor (1953) found that the cloze procedure ranked passages on the basis of difficulty as well as the Flesch and the Dale-Chall readability formulas. A subsequent study showed that the cloze procedure could rank relatively easily worded passages of high concept load, such as those written by Gertrude Stein and James Joyce, better than the formulas mentioned above. Later, studies by Taylor (1957) with the Air Force confirmed that a cloze procedure based on an every-n^{th} deletion system could effectively contrast relative difficulties of different types of material. Cloze had an advantage over the formulas in that it somehow included the idea density of a passage. It had a disadvantage because it had to be administered directly to students, while the formulas could be used to determine difficulty arithmetically.

Bormuth's study (1962, 1963) with intermediate level students further confirmed that "cloze tests were valid and highly reliable predictors of the comprehension difficulties of the passage and appropriate for use with individuals and groups which vary widely in comprehension ability." (Bormuth, 1963 p. 134) Gallant's results (1964, 1965) added further evidence to the validity and reliability using primary level students.

Other efforts have been made using cloze in readability research. Miller and Coleman (1967) used the cloze procedure to calibrate a series of passages ranging in difficulty from easy to complex. They felt that such a scale would be useful in research investigations. Three variations of the

cloze procedure were used: (1) every-fifth word deletions, (2) one deletion per passage; and (3) unilateral guessing, that is, covering all the words in a passage and having the subject guess each word in sequential order. All three methods ranked the passages in approximately the same order. Aquino (1969) conducted a study to validate the Miller-Coleman Readability Scale, mentioned above, with a small, selective sample of subjects. Different forms of cloze tests were used on the same passage, but all methods ranked the passages in approximately the same order. The author concluded that "the economy of the cloze procedure suggests it to be a suitable ranking device for determining passage difficulty." (p. 347)

In 1964, Bormuth (1964a) began using multiple forms of cloze tests over the same passage. By using an every-fifth word deletion system and rotating the starting points, he could make five cloze tests over a given passage and thus get a cloze measure on every word in the passage. He examined "word depth," a concept developed by Yngve (1962) for use in programming computers to translate language, as a measure of the grammatical complexity of sentences. (Bormuth, 1964b) In comparing mean word depth with the Dale-Chall formula, Bormuth found that both methods ranked passages in the same order when the subject matter of the passages was held constant. However, when the Dale-Chall level was held constant but the subject matter allowed to vary, mean word depth was a more powerful predictor. This suggests that mean word depth may be a more effective means of measuring difficulty between content areas. These preliminary studies led to Bormuth's major work in readability (Bormuth, 1966).

This was a multi-purpose study designed to examine several aspects of readability. Bormuth was interested in: (1) determining the forms and

strengths of the relationships between certain measurable language variables and comprehension difficulty; (2) determining if the effects of such variables vary with the reading ability of subjects; (3) determining the possibility of testing the readability of small language units, such as words and phrases; and (4) measuring the accuracy of "new" linguistic variables in predicting language difficulty. Twenty cloze tests were constructed over a variety of content materials and prose styles, and then administered to a large group of fourth through eighth grade students. The results have several important implications. Bormuth found that several variables have curvilinear relationships with comprehension difficulty. Most readability research in the past had assumed linear relationships. The language variables that were measured predicted difficulty equally well for all students, regardless of differences in reading ability. It was found that useful predictions could be made for individual words, independent clauses, and sentences. Precision was increased considerably by the use of linguistic variables such as word depth, letter redundancy, and others. Previous readability formulas had had correlations ranging from .50 to .70 between the variables measured, often sentence length and number of "hard" words, and comprehension difficulty. Bormuth arrived at a .93 correlation between the linguistic variables he used and comprehension difficulty at the passage level. He attributed the significant increase in prediction to two factors: (1) cloze tests were more reliable and valid than the criterion tests that are usually used; and (2) the new language variables that were used were based on linguistic research. Although this study made significant contributions to our knowledge of readability, it is also important for another reason. It signaled the arrival of cloze as an important research tool in the study of language. In the earlier

research efforts, cloze had been used to rank passages. Now cloze offered a means of examining the importance of particular language variables with greatly improved accuracy.

Commenting on the recent research on readability, Bormuth (1968b) has indicated that future efforts at measuring the difficulty of language must be more complex than the oversimplified measures of the past. He predicts that as linguistics adds to our knowledge of language variables, readability formulas "will probably be designed to provide a profile of the level of difficulty represented by each of the language features in a passage." (p. 492) Such formulas would be very complex, probably calculated by computer, and not computed by the classroom teacher.

In the meantime, Bormuth (1967c; 1968c) has outlined specific guidelines for the construction and application of the "cloze readability procedure." In constructing a cloze readability test he suggests the following procedures:

1. Delete every-fifth word in the passage.
2. Replace the deletions with underlined blanks of standardized length.
3. Administer the test to subjects who have never read the passage.
4. Instruct the subjects to fill in the blanks with what they think the deleted words are.
5. Score exact replacements correct.
6. The difficulty level of the passage will be the mean of all the subjects' percentage scores.
7. (Optional) To measure the difficulty of every word, phrase, clause, or sentence, use five cloze forms by rotating the starting points. The percentages of subjects correctly answering each blank is the measure of difficulty of that unit. The difficulty of phrases, clauses, and sentences can be found by averaging the difficulty measure of words within them.

The cloze readability procedure can be used in placing students in graded materials of the appropriate difficulty, in constructing a test that

represents the difficulty of a text, and in selecting materials for a group of students.

In summary, the research has indicated that the cloze procedure can be used effectively to rank passages according to difficulty for subjects with widely varying reading ability. The cloze readability procedure has the advantage over commonly used formulas of being sensitive to the conceptual difficulty of the material. The cloze procedure has also proven successful as a tool for measuring the difficulty of smaller language units, such as words and sentences. A disadvantage of the cloze procedure in determining readability level is that it must be administered directly to the subjects. It is important to remember that all efforts at using cloze for readability purposes have considered only the comprehensibility of the material. Factors such as interest, multiple meanings of words, format, and organization, have not been taken into account. As Bormuth concluded at the end of his study, "there still are no adequate instruments for measuring the interest and esthetic responses that passages elicit in subjects." (Bormuth, 1966 p. 131)

Cloze in the Study of Language Variables

From the efforts that have been made in readability, mentioned in the previous section, have sprung increased interest and concern for particular language variables. Contributions from the science of linguistics have also been vital in stimulating research in this area. The following studies illustrate the usefulness of cloze as a tool for examining specific language variables and their relationship to reading.

The effect of sentence length upon comprehension has been studied by many investigators. Many readability formulas assume that sentence length

is an important factor in the difficulty of language, that is, as sentences get longer, the material gets more difficult to understand. Coleman (1962) used the cloze procedure to measure the comprehension effects of shortening sentences by dividing the compound and complex sentences. He selected three passages of highly technical material and adjusted each so that one had 15.4 words per sentence, another 23.2, and the third 38.7. Cloze tests, based on every-fifth word deletions, were made over each. Subjects were given 50 seconds to read the passage before taking the cloze test. Although the mean number of correct cloze responses increased with a decrease in sentence length, the differences were very small. In analyzing his results, Coleman hypothesized that dividing a sentence joined by "and" does not improve comprehension. He suggested exploring the effects of raising clause fragments, such as subordinate clauses, to full sentences. Unfortunately, Coleman's sample population, 90 undergraduates at John Hopkins University, and the type of material used, do not make his results very generalizable.

In the second phase of her study, Gallant (1964, 1965) prepared comparable sets of passages by controlling for the number of words per passage and the number of "hard" words (those outside the Dale List of 769 Easy Words) but allowing sentence length to vary. Cloze test results indicated that passages with longer sentences were significantly more difficult for first and second graders. However, there was no significant difference in comprehension, due to varied sentence length, for the third graders in her study.

After employing new linguistic variables in his readability study, Bormuth (1966) concluded that the length and complexity of a sentence

could be measured separately and that each correlates differently with comprehension difficulty.

Certainly the results of recent linguistic research and studies like those mentioned above, are causing us to think seriously about the effect of sentence length on comprehension. The cloze procedure is playing an important role in the examination of this problem.

In other studies of language, Ruddell (1965a; 1965b) investigated the effect of the similarity of oral and written patterns of language structure on reading comprehension of fourth grade children. Following up the work of Strickland, he constructed six passages--three using high frequency patterns of oral language and three using low frequency patterns. The cloze procedure was used to measure the students' comprehension of the passages. The results showed that students comprehended materials using high frequency patterns of oral language significantly better than they comprehended passages using low frequency patterns. A similar follow-up study examined the structural elements in the high and low frequency patterns and found reading comprehension to be a function of the redundancy of the syntactical elements used in the materials.

Taylor used the cloze procedure as a measure of entropy. (1956) In information theory, entropy is an estimate of the uncertainty of a situation. For example, if a given cloze item can be correctly identified by a high percentage of subjects, it is considered to have low entropy. If only a few subjects identify the item, it is of high entropy. Taylor found that "cloze scores are dependable estimates of negative entropy . . . or redundancy." (p. 48)

Louthan (1965) used seventh grade students to determine the contributions certain kinds of words make to the meaning of a passage. By using a

variety of cloze tests based on selective deletions of particular kinds of words and followed by comprehension questions, he was able to assess the effect different parts of speech had upon understanding the material. From the results, it was apparent that the greatest loss of comprehension came from deleting nouns, verbs, and modifiers--the basic meaning carriers of the written material.

Blumenfield and Miller (1966) conducted a study to determine what "good" English students knew grammatically that enabled them to learn material more efficiently than poor students. Cloze tests were constructed using an every-fifth word deletion rate and starting points were rotated to obtain measures on all words. Responses were categorized and analyzed according to word class. It was found that there was no significant difference in performance between good and poor students on any word class. On the easier word classes, all students had 100 per cent completion. It appeared from the results that these college students, both good and poor, knew the structure of their language.

Weaver and Bickley (1967) used Rankin's dichotomy of structural-lexical elements to examine the relationship of the written production of language and reading. Two groups of college sophomores wrote stories in response to pictures on the Thematic Apperception Test. These stories were then read by two additional groups. Two days later, all subjects received cloze tests over the stories. Some of the tests were based on lexical deletions, others on structural deletions. Writers reproduced their own lexical items to a greater extent than readers; however, readers produced the writers' structural items as well as the writers themselves. "This implies that in the case of structure, writers and readers of the

language possess identical language elements and highly similar probabilities of the occurence of those language elements in particular context." (p. 290)

From the preceding studies, it can be seen that the cloze procedure is an effective tool for studying aspects of language related to reading. Cloze has been used to investigate entropy or redundancy, sentence length, word depth, and the effects of high and low frequency patterns. Perhaps the greatest usefulness of the cloze procedure is in the study of the structure of language. As Weaver has suggested, "we are on our way toward developing differentials between the syntactic and the semantic, using the cloze procedure. . . if finally realized, this would be a major methodological advance." (Weaver, 1965 p. 131)

Cloze As A Teaching Device

A number of writers and researchers have recommended the cloze procedure as a suitable device for teaching comprehension. Their recommendations are based on the assumption that by going through the task of completing cloze units, a subject will gain insights into the process of using context, recognizing the interrelationships of language, and consequently improve comprehension skills. Very little research has been conducted using the cloze procedure as a teaching technique.

Schneyer (1965) used cloze exercises, based on basal reader material, for students at the sixth grade level. Two types of cloze exercises were used--one based on every-tenth word deletions and the other on noun-verb deletions. The cloze exercises were scored the same day by the teacher, using exact replacement responses, and returned with the correct answers

to the students. The experimental group proceeded with the regular reading program but, in addition, received one cloze exercise a day, alternating between the two types. The control group received the regular reading instruction. At the conclusion of the study, both groups had improved in comprehension, but there was no significant difference for the group using cloze exercises. Analyzing the results, Schneyer points out that "students whose word recognition ability was at the sixth reader level or above performed significantly better on the cloze exercises than did students whose word recognition ability was at fifth reader level or below." (p. 177) Certainly researchers using cloze as a teaching technique would want to make certain that the passages were written at a reading level that was in line with the word recognition ability of their subjects. Schneyer hypothesized that discussion of the reasons for selecting responses might be more effective than just checking for correctness.

Roossinck (1962) also used the cloze procedure in teaching sixth grade students. She developed a type of programmed learning procedure which consisted of a series of 200 cloze exercises graduated in difficulty. Students received immediate feedback after completing a cloze item. Credit was given for synonyms.

Heitzman and Bloomer (1967; Bloomer, 1966) hypothesized that the act of filling in a cloze unit was in itself intrinsically reinforcing for the subject--a non-overt reinforcement. In the first phase of their study, fifth, seventh, ninth, and eleventh graders received cloze exercises for three weeks. In the second phase, termed "longitudinal," ninth graders continued working two cloze exercises per week for a period of twelve weeks. Subjects were randomly assigned to one of seven groups. One group

served as a control, reading the passages intact and answering the comprehension questions that followed. Other groups worded the cloze exercises, some based on every-tenth word deletions, others on noun deletions, verb deletions, and the like, answering comprehension questions after each paragraph. The Iowa Test of Basic Skills was used as the criterion measure. The results indicated that "the use of non-overt reinforced cloze procedure does not increase reading ability either during the process or as a function of post-treatment testing." (Heitzman and Bloomer, 1967 p. 218) The cloze precedure was no better than the conventional reading comprehension exercises. The authors felt that the value of cloze in teaching comprehension is directly related to the method by which it is delivered. Their suggestions for increasing the effectiveness of cloze in teaching included: (1) more reinforcement by the teacher for correct responses, including synonyms; and (2) providing a motivational scheme in that a subject's movement through the exercises is contingent upon the quality of his responses.

Bloomer also used the cloze procedure as a remedial teaching technique for college students. (Bloomer, 1962) Of the three matched groups in the study, one received cloze exercises based on every-tenth word deletions, a second proceeded with traditional remedial exercises, and the third group received no treatment at all. Pre and post testing with the Diagnostic Survey Test revealed that the group using cloze exercises increased significantly more in comprehension and total reading ability. Also, the achieved grade point averages for the cloze group were greater than predicted grade point averages made at the beginning of the study. As a result of this study, Bloomer felt "the cloze procedure does have a positive

effect on comprehension and college grades." (p. 178) However, experimental mortality, regression effects, and lack of adequate control make one skeptical of the results of this study.

Guice (1969) used cloze exercises in addition to regular instruction in reading comprehension for an experimental group of college students. The control group received just the regular instruction in comprehension. Cloze exercises were based on every-tenth deletions of concept words (nouns, verbs, adjectives or adverbs). Two points were scored for exact replacement and one point for synonyms. On the basis of pre and post testing on the reading comprehension section of the Cooperative English Test, the experimental group did not improve significantly more than the control group. It appeared from the results that other factors were at play in Guice's study. That is, both afternoon groups, regardless of treatment, did better than the morning groups.

In the second phase of their study, Blumenfield and Miller (1966) tried to implement the findings of the first phase. Their emphasis was on using the cloze procedure to teach college students grammatical concepts of language. Pre and post testing with the Davis Reading Test showed no significant difference for those using cloze exercises. The authors suggested that perhaps the effectiveness of cloze as a teaching device is related to the type of deletions made, i.e. every-n^{th}, nouns, verbs, and the like.

Friedman (1964) employed the cloze procedure in teaching foreign students. She constructed 20 cloze exercises, using every-fifth word deletions, over materials from McCall Crabbs Standard Test Lessons in Reading. An experimental group received two cloze exercises per week for ten weeks. Credit was given for synonyms, and multiple-choice comprehension questions

followed each cloze exercise. By contrast, the control group received four regular McCall-Crabbs lessons per week for ten weeks. Although both groups made gains in comprehension, there was no significant difference between the mean gains of the two groups. However, the difference in the amount of instruction the two groups received may have had an effect.

Contrary to the recommendations frequently made by authorities in the field, the research evidence at the present time does not suggest the cloze procedure as an effective teaching technique. Independent studies using a range of age levels have demonstrated that the cloze procedure, used either as a supplement to or in lieu of "regular" reading instruction, does not produce significantly improved results in comprehension. Lest we dismiss cloze altogether as a teaching technique, future research should explore alternative procedures. Culhane (1970) suggests making the blanks the length of the deleted words, thereby adding another clue. Discussions following cloze exercises could give students opportunities to express why they chose particular responses. Future efforts might also focus on the effects various types of deletions have upon teaching comprehension. Louthan (1965) has suggested that perhaps cloze may be used as a technique to convey an understanding of what kinds of words bear the burden of communication in written material. Clearly, there is the need for more research on the effectiveness of the cloze procedure as a teaching device.

Summary and Conclusions

This paper has surveyed the literature related to the cloze procedure. Research in the areas of methodology, the use of cloze as a measuring device, and the effectiveness of cloze as a teaching technique has been considered. Cloze tests based on mechanical deletions of every-n^{th}, usually

every-fifth, have been recommended most often. Other deletion systems have also been used, such as random deletions of a certain percentage of words and selective deletions based on form class. Scoring exact replacements is the most efficient and useful procedure used. Test length of 250 words with a minimum of 50 deletions has often been suggested but some investigators have used much longer tests. It seems that the cloze procedure is a valid and reliable measure of comprehension ability, both for children and adults. Cloze test results have correlated highly with multiple-choice comprehension questions over the same passage and with standardized reading tests. Other investigators have used the cloze procedure to explore aspects related to comprehension, such as literary style, attitudes toward the content, and listening comprehension. Cloze has also proven effective as a means of ranking passages according to difficulty. When used in this manner, cloze correlates highly with traditional readability formulas but has the advantage of being sensitive to the conceptual difficulty of the material. In more recent efforts, the cloze procedure has been coupled with the use of linguistic variables for a new approach to the study of readability. Cloze provides for a measure of difficulty of smaller language units, such as words, phrases, and sentences, not possible with traditional readability measures. Because of this, cloze has been valuable in examining the influence of particular language variables upon reading. Sentence length, redundancy, and syntactical structure are just some of the variables that have been studied using the cloze procedure. Limited efforts have been made to assess the efficacy of cloze as a teaching technique. At the present time, the evidence does not favor cloze over more conventional methods of teaching comprehension.

In reviewing the literature some general weaknesses become apparent. Some investigators fail to report the kind of cloze procedure being used. They neglect to adequately describe the type, rate, and number of deletions made and the rationale behind the system used. The nature of the content and the difficulty of the material used for the cloze passage are frequently not reported. Scoring systems are often not explicated. In studies where cloze results are correlated with comprehension questions, researchers often neglect to fully describe the comprehension questions, explain how they were developed, or validate them prior to use. Although some researchers briefly mention the grade level of the students used in their studies, complete descriptions of subjects are woefully inadequate. A common and serious omission is the failure fo report the reading levels of the subjects. External validity has been seriously hampered in many studies because of the highly select samples of subjects--often taken from a class the investigator is teaching at the time of the study. In studies investigating cloze as a teaching technique by comparing it with other methods, the "regular" method of instruction is seldom defined. In light of the weaknesses that have been noted in the literature, the following guidelines are suggested for the reporting of cloze research:

1. Report the type of cloze procedure used. This should include the type of deletions, the rate of deletions, the total number of deletions, and the rationale for using this particular system.

2. Describe the material upon which the cloze passages are based. This should include the type of content, the style, and an estimate of the readability level.

3. Explain the scoring system used.

4. Fully describe the subjects used in the study. Descriptions should be based on factors such as reading level, grade placement, sex, socio-economic status, geographical environment, as well as other relevant variables.

5. If cloze results are to be correlated with other comprehension measures, describe those measures, state the types of comprehension questions used, and, if possible, prevalidate them before use in the study.

6. If the cloze procedure is being compared to other teaching methods, describe such methods in terms of materials, time spent in instruction, and philosophy behind the program.

A number of opportunities exist for future research using the cloze procedure. Very little is known about the construct validity of the cloze procedure as a measure of comprehension ability. Factor analysis studies might aid at getting at the underlying processes involved in cloze. Perhaps the Campbell-Fiske convergent and discriminant validity paradigm might be a useful approach. As Jenkinson has shown, the cloze procedure may be used to examine the components of comprehension itself. Future efforts in the study of the concept of information gain should be encouraged. The results of such studies should lead to a re-evaluation of traditional standards of instructional difficulty and give us new information as to the level of passage difficulty that produces the greatest gain in student knowledge.

Future research in readability should parallel and build upon new developments in linguistic and psycholinguistic theory. As we discover more knowledge about various aspects of language, we should increase our understanding of just what makes written language difficult to understand. The cloze procedure could be combined with measures of affect and attitude, such as a semantic differential, to get at aspects of readability that are often neglected. Perhaps, as we learn to control language, in the not-too-distant future, we will have developed style guides to assist writers of instructional materials. Cloze can play an important role in future readability research.

Perhaps the greatest contribution the cloze procedure could make in future research is in the study of the structure of language and its influence

on reading. For example, the study done by Blumenfield and Miller (1966) could be replicated but applied to elementary level students to determine at what age students acquire an adequate knowledge of grammar. The results of such a study could have important implications for the teaching of reading and language.

New approaches need to be tried in using cloze as a teaching technique. Rather than attempting to determine the superiority of cloze over other teaching methods, future efforts might better be directed towards determining what benefits cloze offers as a teaching device and then using cloze to supplement conventional methods. One of the reasons cloze has not fared too well as a teaching technique in past research might be that investigators have frequently relied on the technique itself to do the teaching. New approaches should definitely take this into consideration.

In addition to the uses of cloze reviewed in this paper, other untried research possibilities exist. In an Occasional Report for the Center for the Study of Evaluation of Instructional Programs, Bormuth (1967d) suggests the use of cloze in constructing criterion-referenced tests for evaluating instructional programs. Rankin (1959a) suggests a number of clinical uses for the cloze procedure, such as measuring improvement due to remediation, providing greater transfer between the classroom and the remedial situation, as well as others. Taylor (Greene, 1968) has suggested exploring other scoring procedures such as latency and gambling instinct, for specific purposes. A number of areas remain to be explored.

Although research with the cloze procedure has contributed a great deal to our knowledge of reading and language, much more remains to be investigated before we can fully judge the effectiveness of cloze as a measurement and teaching technique.

REFERENCES

Aquino, Milagros R. The Validity of the Miller-Coleman readability scale, Reading Research Quarterly, 4 (Spring 1969) 342-57.

Bloomer, Richard H. The cloze procedure as a remedial reading exercise. Journal of Developmental Reading, 5 (Spring 1962) 173-81.

Bloomer, Richard et. al. Nonovert reinforced cloze procedure. U. S. Office of Education, Cooperative Research Project #2245, University of Connecticut, 1966, 152 p. (ERIC-ED010-270).

Blumenfield, Jacob P. and Gerald R. Miller. Improving reading through teaching grammatical constraints. Elementary English, 43 (November 1966) 752-55.

Bormuth, John. Cloze tests as measures of readability and comprehension ability. Unpublished doctoral dissertation, Indiana University, 1962.

Bormuth, John. Cloze as a measure of readability. Proceedings of the International Reading Association, 1963, 8, 131-34.

Bormuth, John. Experimental application of cloze tests. Proceedings of the International Reading Association, 1964, 9, 303-06.(a)

Bormuth, John. Mean word depth as a predictor of comprehension difficulty. California Journal of Educational Research, 15 (November 1964) 226-31. (b)

Bormuth, John. Relationships between selected language variables and comprehension ability and difficulty. U. S. Office of Education Cooperative Research Project #2082, 1964, (ERIC-ED003-325). (c)

Bormuth, John. Validities of grammatical and semantic classifications of cloze test scores. Proceedings of the International Reading Association, 1965, 10, 284-85. (a)

Bormuth, John. Optimum sample size and cloze test length in readability. Journal of Educational Measurement, 2 (June 1965) 111-16. (b)

Bormuth, John. Readability: a new approach. Reading Research Quarterly, 1 (Spring 1966) 79-135.

Bormuth, John. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 10 (February 1967) 291-300. (a)

Bormuth, John. Design of readability research. Proceedings of the International Reading Association, 1967, 11, part 1, 485-89. (b)

Bormuth, John. Cloze readability procedure. Occasional Report #1, Center for the Study of Evaluation of Instructional Programs, California University: Los Angeles, (February 1967) 29 p. (ERIC-ED010-983). (c)

Bormuth, John. The implications and use of cloze procedure in the evaluation of instructional programs. Occasional Report #3, Center for the Study of Evaluation of Instructional Programs, California University: Los Angeles, (April 1967) 55 p. (ERIC-ED012-674). (d)

Bormuth, John. Cloze test reliability: criterion reference scores. Journal of Educational Measurement, 5 (Fall 1968) 189-96. (a)

Bormuth, John. New data on readability. Proceedings of the International Reading Association, 1968, 12, part 1, 488-92. (b)

Bormuth, John. The cloze readability procedure. Elementary English, 45 (April 1968) 429-36. (c)

Bormuth, John. Factor validity of cloze tests as measures of reading comprehension ability. Reading Research Quarterly, 4 (Spring 1969) 358-65. (a)

Bormuth, John. Empirical determination of the instructional reading level. Proceedings of the International Reading Association, 1969, 13, part 1, 716-21. (b)

Bormuth, John and C. L. MacDonald. Cloze tests as a measure of ability to detect literary style. Proceedings of the International Reading Association, 1965, 10, 287-90.

Chall, Jeanne. Readability: an appraisal of research and application. Bureau of Educational Research Monographs #34, Columbus, Ohio: Ohio State University, 1958, 202 p.

Coleman, E. B. Improving comprehensibility by shortening sentences. Journal of Applied Psychology, 46 (1962) 131-34.

Coleman, E. B. and G. R. Miller. A measure of information gained during prose learning. Reading Research Quarterly, 3 (Spring 1968) 369-86.

Culhane, Joseph W. Cloze procedures and comprehension. Reading Teacher, 23 (February 1970) 410-13.

Fillenbaum, Samuel and Lyle V. Jones and Amnon Rapoport. The predictability of words and their grammatical classes as a function of rate of deletion from a speech transcript. Journal of Verbal Learning and Verbal Behavior, 2 (August 1963) 186-94.

Friedman, M. The use of the cloze procedure for improving the reading of foreign students at the University of Florida. Unpublished doctoral dissertation, University of Florida, 1964.

Gallant, Ruth. An investigation of the use of cloze tests as a measure of readability of materials for the primary grades. Unpublished doctoral dissertation, Indiana University, 1964.

Gallant, Ruth. Use of cloze tests as a measure of readability in the primary grades. Proceedings of the International Reading Association, 1965, 10, 286-87.

Greene, Frank. Modification of the cloze procedure and changes in reading test performances. Journal of Educational Measurement, 2 (1965) 213-17.

Greene, Frank, et. al. Cloze symposium. Seventeenth Yearbook of the National Reading Conference, Milwaukee, Wisconsin: National Reading Conference, 1968, 110-31.

Guice, Billy M. The use of the cloze procedure for improving reading comprehension of college students. Journal of Reading Behavior, 1 (Summer 1969) 81-92.

Hafner, L. E. Relationships of various measures to the cloze. Thirteenth Yearbook of the National Reading Conference, Milwaukee, Wisconsin: National Reading Conference, 1964, 135-45.

Hafner, L. E. Implication of cloze. Fourteenth Yearbook of the National Reading Conference, Milwaukee, Wisconsin: National Reading Conference, 1965, 151-58.

Hafner, L. E. Cloze procedure. In "Research for the Classroom" Ed. by A. J. Kingston, Journal of Reading, 9 (May 1966) 415-21.

Heitzman, A. J. and R. Bloomer. The effect of non-overt reinforced cloze procedure upon reading comprehension. Journal of Reading, 11 (December 1967) 213-223.

Jenkinson, Marion. Selected processes and difficulties of reading comprehension. Unpublished doctoral dissertation, University of Chicago, 1957.

Kingston, A. J. and W. W. Weaver. Recent developments in readability appraisal. Journal of Reading, 11 (October 1967) 34-47.

Klare, George R. The Measurement of Readability, Ames, Iowa: Iowa State University Press, 1963, 328 p.

Klare, George R. Comments on Bormuth's readability: a new approach. Reading Research Quarterly, 1 (Summer 1966) 119-25.

Louthan, Vincent. Some systematic grammatical deletions and their effects on reading comprehension. English Journal, 54 (April 1965) 295-99.

Manis, Melvin and Robyn M. Dawes. Cloze score as a function of attitude. Psychological Reports, 9 (1961) 79-84.

MacGinitie, Walter H. Contextual constraint in English prose paragraphs. Journal of Psychology, 51 (1961) 121-30.

McLeod, J. and J. Anderson. Readability assessment and word redundancy of printed English. Psychological Reports, 18 (February 1966) 35-38.

Miller, G. R. and E. B. Coleman. A set of thirty-six prose passages calibrated for complexity. Journal of Verbal Learning and Verbal Behavior, 6 (December 1967) 851-54.

Mills, Robert E. and Jean R. Richardson. What do publishers mean by grade level? Reading Teacher, 16 (March 1964) 359-62.

Musgrave, Barbara. The effect of who and what contexts on cloze and commonality scores. Journal of Social Psychology, 59 (1963) 185-92.

Rankin, Earl. An evaluation of the cloze procedure as a technique for measuring reading comprehension. Unpublished doctoral dissertation, University of Michigan, 1957.

Rankin, Earl. Uses of the cloze procedure in the reading clinic. Proceedings of the International Reading Association, 1959, 4, 228-32. (a)

Rankin, Earl. The cloze procedure--its validity and utility. Eighth Yearbook of the National Reading Conference, Milwaukee, Wisconsin: National Reading Conference, 1959, 131-44. (b)

Rankin, Earl. The cloze procedure--a survey of research. Fourteenth Yearbook of the National Reading Conference, Milwaukee, Wisconsin: National Reading Conference, 1964, 133-50.

Rankin, Earl. Research design and the cloze procedure. Proceedings of the International Reading Association, 1967, 11, part 1, 489-91.

Ransom, Peggy E. Determining reading levels of elementary school children by cloze testing. Proceedings of the International Reading Association, 1968, 12, part 1, 477-82.

Roossinck, P. L. A method for improving reading comprehension. Unpublished master's thesis, University of Michigan, 1962.

Ruddell, Robert. A study of the cloze comprehension technique in relation to structurally controlled reading material. Proceedings of the International Reading Association, 1964, 9, 298-302.

Ruddell, Robert. The effect of oral and written patterns of language structure on reading comprehension. Reading Teacher 18 (January 1965) 270-75. (a)

Ruddell, Robert. Reading comprehension and structural redundancy in written material. Proceedings of the International Reading Association, 1965, 10, 308-11. (b)

Schneyer, J. Wesley. Use of the cloze procedure for improving reading comprehension. Reading Teacher, 19 (December 1965) 174-79.

Taylor, Wilson L. Cloze procedure: a new tool for measuring readability. Journalism Quarterly, 30 (Fall 1953) 415-33.

Taylor, Wilson L. Recent developments in the use of cloze procedure. Journalism Quarterly, 33 (Winter 1956) 42-48.

Taylor Wilson L. Cloze readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology, 41 (February 1957) 19-26.

Weaver, W. W. The predictability of omissions in reading and listening. Eleventh Yearbook of the National Reading Conference, Milwaukee, Wisconsin: National Reading Conference, 1962, 148-53.

Weaver, W. W. Theoretical aspects of the cloze procedure. Fourteenth Yearbook of the National Reading Conference, Milwaukee, Wisconsin: National Reading Conference, 1965, 115-132.

Weaver, W. W. and A. C. Bickley. Structural-lexical predictability of materials which predictor has previously produced or read. Conference Proceedings of the 75th Annual Convention, American Psychological Association, 1967, 289-90.

Weaver, W. W. and A. J. Kingston. A factor analysis of the cloze procedure and other measures of reading and language ability. Journal of Communication, 13 (December 1963) 252-261.

Weaver, W. W., William White and A. J. Kingston. Affective correlates of reading comprehension. Journal of Psychology, 68 (1968) 87-95.

Weintraub, Sam. The cloze procedure. Reading Teacher, 21 (March 1968) 567-71.

Yngve, Victor. Computer programs for translation. Scientific American, 206 (June 1962) 68-76.